Section 3

Introduction to Sampling

A *population* consists of all members of a group in which a researcher has an interest. It may be small, such as all psychiatrists affiliated with a particular hospital, or it may be large, such as all high school seniors in a state. When populations are large, researchers usually use a sample. A *sample* is a subset of a population. For instance, we might be interested in the attitudes of all registered nurses in Texas toward people with AIDS. The nurses would constitute the population. If we administered an AIDS attitude scale to all these nurses, we would be studying the population, and the summarized results (such as averages) would be referred to as *parameters.* If we studied only a sample of the nurses, the summarized results would be referred to as *statistics.*

A *population* consists of all members of a group.

A *sample* is a subset of a population.

Populations yield *parameters.*

Samples yield *statistics.*

No matter how a sample is drawn, it is always possible that the *statistics* obtained by studying the sample do not accurately reflect the *population parameters* that would have been obtained if the entire population had been studied. In fact, researchers almost always expect some amount of error as a result of sampling.

If sampling creates errors, why do researchers sample? First, for economic and physical reasons, it is not always possible to study an entire population. Second, with proper sampling, highly reliable results can be obtained. Furthermore, with proper sampling, the amount of error to allow for in the interpretation of the resulting data can be estimated with inferential statistics, which are covered in Part D of this book.

Freedom from *bias* is the most important characteristic of a good sample. A bias exists whenever some members of a population have a greater chance of being selected for inclusion in a sample than other members of the population. Here are three examples of biased samples:

A *bias* exists whenever some members of a population have a greater chance of being selected than others.

- A professor wishes to study the attitudes of all sophomores at a college (the population) but asks only those enrolled in her introductory psychology class (the sample) to participate in

the study. Note that only those in the class have a chance of being selected; other sophomores have no chance.

- An individual wants to predict the results of a citywide election (the population) but asks the intentions of only voters whom he encounters in a large shopping mall (the sample). Note that only those in the mall have a chance of being selected; other voters have no chance.

- A magazine editor wants to determine the opinions of all rifle owners (the population) on a gun-control measure but mails questionnaires only to those who subscribe to her magazine (the sample). Note that only subscribers to her magazine have a chance to respond; other rifle owners have no chance.

Samples of convenience are biased.

In the above examples, *samples of convenience* (or *accidental samples*) were used, increasing the odds that some members of a population would be selected while reducing the odds that other members would be selected. In addition to the obvious bias in the examples, there is an additional problem. Even those who do have a chance of being included in the samples may refuse to participate. This problem is often referred to as the problem of *volunteerism* (also called *self-selection bias*). Volunteerism is presumed to create an additional source of bias because those who decide not to participate have no chance of being included. Furthermore, many studies comparing participants (i.e., volunteers) with nonparticipants suggest that participants tend to be more highly educated and tend to come from higher socioeconomic status (SES) groups than do their counterparts. Efforts to reduce the effects of volunteerism include offering rewards, stressing to potential participants the importance of the study, and making it easy for individuals to respond, such as by providing them with a stamped, self-addressed envelope.

Volunteerism in sampling is presumed to create a bias.

To eliminate bias in the selection of individuals for a study, some type of *random sampling* is needed. A classic type of random sampling is *simple random sampling*. This technique gives each member of a population an equal chance of being selected. A simple way to accomplish this with a small population is to put the names of all members of a population on slips of paper, thoroughly mix the slips, and have a blindfolded assistant select the number desired for the sample.[1] After the names have been selected, efforts must be made to encourage all those selected to participate. If some refuse, as

In *simple random sampling*, each member of a population is given an equal chance of being selected.

often happens, a biased sample is obtained even though all members of the population had an equal chance to have their names selected.

Suppose that a researcher is fortunate. He or she selected names using simple random sampling and obtained the cooperation of everyone selected. In this case, the researcher has obtained an *unbiased sample.* Can the researcher be certain that the results obtained from the sample accurately reflect those results that would have been obtained by studying the entire population? Certainly not. The possibility of random errors still exists. Random errors (created by random selection) are called *sampling errors* by statisticians. At random (i.e., by chance), the researcher may have selected a disproportionately large number of Democrats, males, low SES group members, and so on. Such errors make the sample unrepresentative and therefore may lead to incorrect results.

Simple random sampling identifies an *unbiased sample.*

Random sampling produces *sampling errors.*

If both biased and unbiased sampling are subject to error, why do researchers prefer unbiased random sampling? They prefer it for two reasons: (1) inferential statistics, which are described in Part D of this book, enable researchers to estimate the amount of error to allow for when analyzing the results from unbiased samples, and (2) the amount of sampling error obtained from unbiased samples tends to be small when large samples are used.

While using large samples helps to limit the amount of random error, it is important to note that selecting a large sample does not correct for errors due to bias. For instance, if the individual trying to predict the results of a citywide election in the second example above is very persistent and spends weeks at the shopping mall asking shoppers how they intend to vote, the individual will obtain a very large sample of people who may differ from the population of voters in various ways, such as by being more affluent, having more time to spend shopping, and so on. This illustrates that increasing the size of a biased sample does not reduce the amount of error due to bias.

Selecting a large sample does not correct for errors due to bias.

Despite the above discussion, it is *not* true that all research in which biased samples are used is worthless. There are many situations in which researchers have no choice but to use biased samples. For instance, for ethical and legal reasons, much medical research is conducted using volunteers who are willing to risk taking a new medication or undergoing a new surgical procedure. If promising results are obtained in initial studies, larger studies with better (but usually still biased) samples are undertaken. At some point, despite the

Often, researchers have no choice but to use biased samples.

possible role of bias, decisions, such as Food and Drug Administration approval of a new drug, need to be made on the basis of data obtained with biased samples. Little progress would be made in most fields if the results of all studies with biased samples were summarily dismissed.

At the same time, it is important to note that the statistical remedies for errors due to biased samples are extremely limited. Because researchers usually do not know the extent to which a particular bias has affected their results (e.g., they do not know how nonrespondents to a questionnaire would have answered the questions), it is generally not possible to adjust statistically for errors created by bias. Thus, when biased samples are used, the results of statistical analyses of the data should be viewed with great caution.

Statistical results based on observations of biased samples should be viewed with great caution.

Various methods of random sampling are described in more detail in Section 17 of this book. Considerations in determining sample size are discussed in Section 18.

Endnote

[1] Another method for selecting a *simple random sample*, and other types of random samples, is described in Section 17.

Exercise for Section 3

Factual Questions

1. What term is used to refer to all members of a group in which a researcher has an interest?
2. If samples yield "statistics," what do populations yield?
3. What is the most important characteristic of a good sample?
4. If a researcher uses a sample of volunteers from a population, should we presume that the sample is biased?
5. What type of sampling eliminates bias in the selection of participants?
6. Briefly describe how one could select a simple random sample.